

Exploring the Manifold of Seismic Waves: Application to the Estimation of Arrival-Times

K.M. Taylor^a, M.J. Procopio^b, C.J. Young^b, F.G. Meyer^{c,*}

^a*Department of Applied Mathematics, University of Colorado at Boulder, Boulder, CO*

^b*Sandia National Laboratories, Albuquerque, NM*

^c*Department of Electrical Engineering, University of Colorado at Boulder, Boulder, CO*

Abstract

We propose a new method to analyze seismic time series and estimate the arrival-times of seismic waves. Our approach combines two ingredients: the times series are first lifted into a high-dimensional space using time-delay embedding; the resulting phase space is then parametrized using a nonlinear method based on the eigenvectors of the graph Laplacian. We validate our approach using a dataset of seismic events that occurred in Idaho, Montana, Wyoming, and Utah, between 2005 and 2006. Our approach outperforms methods based on singular-spectrum analysis, wavelet analysis, and STA/LTA.

Keywords: Time series analysis, eigenvectors of the graph Laplacian, Time-delay embedding, Statistical seismology

1. Introduction

1.1. Estimation of arrival-times

Seismic waves come in distinct bursts, or arrivals, corresponding to different propagation paths through the earth. Arrival-times of seismic waves are indispensable to the determination of the location and type of the seismic event; the precise estimation of arrival-times remains therefore a fundamental problem. This paper addresses the problem of estimating arrival-times of local seismic waves from a seismogram. Several methods for estimating arrival-times use some variants of the classic current-value-to-predicted-value ratio method (e.g. [1, 2, 3] and references therein). The current value is a short term average (STA) of the energy of the incoming data, while the predicted value is a long term average (LTA), so the ratio is expressed as STA/LTA. This ratio is constantly updated as new data flows in, and a detection is declared when the ratio exceeds a threshold value. When the signal and the noise are Gaussian distributed, the STA/LTA method yields an optimal detector that strikes the optimal balance between Type I and Type II errors [4, 5]. As explained in [6], seismic waves are non-Gaussian, and therefore higher order statistics (such as skewness and kurtosis) can be used to detect the onset of seismic waves [7, 8, 9]. The performance of a detector can be improved by enhancing the signal transients relative to the background noise. Several time-frequency and time-scale decompositions have been proposed for this purpose (e.g. [10, 11, 12], and references therein). Advanced statistical methods can use training data (in the form of seismograms labelled by an analyst). For instance, the software developed at the Prototype International Data Center (Arlington, VA) is based on a multi-layer neural network that uses labelled waveforms in order to predict the types of waves of unseen seismograms [13].

*Corresponding author

Email address: E-mail: fmeyer@colorado.edu (F.G. Meyer)

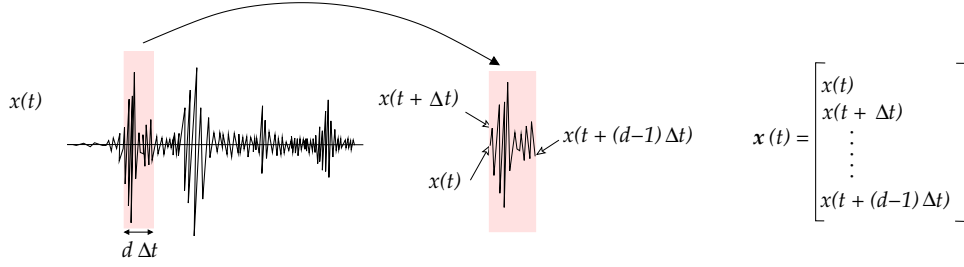


Figure 1: For each time t we construct a patch $\mathbf{x}(t)$ composed of d equally-spaced samples of the seismogram.

1.2. Local analysis of a time series: the concept of patch

In order to detect the arrival of a seismic wave, and estimate its arrival-time, we propose to characterize the local dynamics of the seismogram using a sliding window, or temporal *patch* (see Fig. 1). Formally, let $x(t)$ be one of the components (e.g. the z component, or the first eigenmode after a singular value decomposition of the three components) of a seismogram. Let Δt be the sampling period of the seismogram. The *patch* $\mathbf{x}(t)$ is formed by collecting d equally-spaced samples of $x(t)$ and stacking them into a d -dimensional vector,

$$\mathbf{x}(t) = [x(t) \quad x(t + \Delta t) \quad \dots \quad x(t + (d-1)\Delta t)]^T. \quad (1)$$

While we generally think of $\mathbf{x}(t)$ as a snippet of the original signal over the interval $[t, t + d\Delta t]$ (see Fig. 1), in this paper we will think about $\mathbf{x}(t)$ as a vector in d dimensions (see Fig. 2). We call *patch-space* the region of \mathbb{R}^d formed by the patch trajectory $\{\mathbf{x}(t), t \geq 0\}$ (see Fig. 2). Our goal is to detect the presence of a seismic wave in the patch $\mathbf{x}(t)$ from the location of the patch in patch-space.

The concept of patches is equivalent to the concept of *time-delay coordinates* in the context of the analysis of a dynamical system from the time series generated by an observable [14, 15, 16]. In this work, we offer a novel perspective on the concept of time-delay coordinates by combining several patch trajectories (from several seismograms) and computing a nonlinear parametrization of the combined phase spaces defined by the delay coordinates (1). This nonlinear parametrization assigns to each patch $\mathbf{x}(t)$ a small number of coordinates that uniquely characterize the position of the patch in patch space, while providing the optimal separation between baseline patches and arrival patches.

1.3. Problem statement

We are interested in detecting seismic waves and estimating the arrival-time of each wave. We model the seismogram $x(t)$ as a sum of two components

$$x(t) = b(t) + w(t), \quad (2)$$

where $w(t)$ represents a seismic wave arriving at time τ , and $b(t)$ represents the baseline (or background) activity. We assume that $b(t)$ is a slowly varying signal with its energy uniformly distributed over time. In contrast, we expect that $w(t)$ will be a fast oscillatory transient localized around the arrival-time τ (see Fig. 3). Our goal is to detect the seismic wave $w(t)$, and estimate its onset τ . The difficulty of the problem stems from the fact that there is a large variability in the shape and frequency content of the seismic waves $w(t)$.

We tackle this question by lifting the model (2) into \mathbb{R}^d using the time-delay embedding defined by (1). As explained in Sec. 2.2, after time-delay embedding, *baseline patches* that do not overlap with the seismic wave $w(t)$ and only contain the baseline signal $b(t)$ become tightly clustered along low-dimensional curves. In contrast, *arrival patches* that include portions of the seismic wave $w(t)$ remain at a large distance from one another, and are also at a large distance from the baseline patches. The differential organization of the baseline and arrival patches in \mathbb{R}^d is the first ingredient of our approach.

The second ingredient of our approach is provided by a parametrization of patch-space that manages to cluster baseline patches and arrival patches into two separate groups. This parametrization allows us

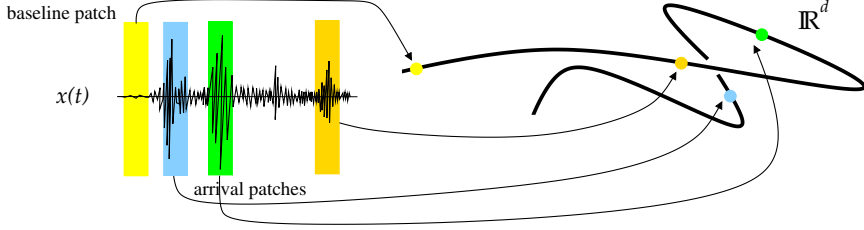


Figure 2: The trajectory $\{\mathbf{x}(t), t \geq 0\}$ is a one-dimensional curve in \mathbb{R}^d . This curve revisits already seen regions of \mathbb{R}^d when the current patch $\mathbf{x}(t)$ is similar to a previously seen patch (e.g. the gold and blue patches).

to represent patches from \mathbb{R}^d , where d is of the order of 10^3 using only about 25 coordinates. Finally, the last stage of our approach consists in training a classifier to detect patches containing seismic waves. The classifier uses the low-dimensional parametrization of patch-space.

In summary, the contribution of this paper is a novel method to analyze seismograms and estimate arrival-times of seismic waves. Our approach includes the following three steps:

1. Construction of patch-space by time-delay embedding of seismograms;
2. Low-dimensional parametrization of patch-space;
3. Construction of a classifier using the low-dimensional parametrization; detection of the presence of seismic waves and estimation of arrival-times.

1.4. Outline

In the next section we explain how the properties of a waveform $x(t)$ will manifest themselves in terms of geometric properties of the patch trajectory $\{\mathbf{x}(t), t \geq 0\}$. In particular, we estimate mutual distances between patches and we describe the alignment of patches along low-dimensional subspaces. Our analysis is performed assuming time is continuous (in Sec. 3.2, we consider the discrete version of the problem). Furthermore, we assume that the set of patches is constructed from a single seismogram. In Sec. 3 we expand our discussion to the case where patches are extracted from several seismograms collected at different stations. In Sec. 3.3, we construct a low-dimensional parametrization of the discrete version of patch-space. We use this low-dimensional parametrization in Sec. 4 to build a classifier that learns to detect patches made up of seismic waves. Finally, the performance of our approach is quantified in Sec. 5.

2. Patch-Space

2.1. Patch trajectories are one-dimensional

Given a seismogram $x(t)$, the patch $\mathbf{x}(t)$ extracted at time t is a vector in \mathbb{R}^d . As t evolves, the set of points $\{\mathbf{x}(t), t \geq 0\}$ forms the patch trajectory. If $x(t)$ is a smooth function of time (has s derivatives), then the patch trajectory is a smooth one-dimensional curve in \mathbb{R}^d , and the dimension of the curve (one) is independent of the patch size d . Indeed, for $k = 0, \dots, d-1$ the maps

$$t \longrightarrow x(t + k\Delta t) \quad (3)$$

are all smooth. Therefore each of the coordinates of $\mathbf{x}(t)$ is a smooth function of t , and the map

$$t \longrightarrow \mathbf{x}(t) \quad (4)$$

is a smooth map from \mathbb{R} to \mathbb{R}^d and thus is a one-dimensional curve in \mathbb{R}^d .

2.2. Mutual Distance Between Two Patches

In the following, we assume that the seismogram is described by the model (2) and we study the Euclidean distance between any two patches $\mathbf{x}(t_i)$ and $\mathbf{x}(t_j)$ extracted at times t_i and t_j . We first consider the case where both patches come from the baseline part of the signal. In this case, we assume $w(t) = 0$ over the intervals $[t_i, t_i + d\Delta t)$ and $[t_j, t_j + d\Delta t)$, and we have

$$\|\mathbf{x}(t_i) - \mathbf{x}(t_j)\|^2 = \sum_{k=0}^{d-1} (b(t_i + k\Delta t) - b(t_j + k\Delta t))^2.$$

If the baseline signal varies slowly, then we have $|b(t_i + k\Delta t) - b(t_j + k\Delta t)| \approx 0$, ($k = 0, \dots, d-1$), and therefore

$$\|\mathbf{x}(t_i) - \mathbf{x}(t_j)\|^2 = \sum_{k=0}^{d-1} (b(t_i + k\Delta t) - b(t_j + k\Delta t))^2 \approx 0.$$

We now consider the case where one patch, $\mathbf{x}(t_i)$ (without loss of generality), is part of a seismic wave $w(t)$, whereas the other patch, $\mathbf{x}(t_j)$, comes from the baseline part. We have

$$\begin{aligned} \|\mathbf{x}(t_i) - \mathbf{x}(t_j)\|^2 &= \sum_{k=0}^{d-1} (w(t_i + k\Delta t) + b(t_i + k\Delta t) - b(t_j + k\Delta t))^2 \\ &= \sum_{k=0}^{d-1} w^2(t_i + k\Delta t) + 2 \sum_{k=0}^{d-1} w(t_i + k\Delta t) (b(t_i + k\Delta t) - b(t_j + k\Delta t)) \\ &\quad + \sum_{k=0}^{d-1} (b(t_i + k\Delta t) - b(t_j + k\Delta t))^2. \end{aligned}$$

Again, we can assume that for each k , $|b(t_i + k\Delta t) - b(t_j + k\Delta t)| \approx 0$, and thus

$$\sum_{k=0}^{d-1} w(t_i + k\Delta t) (b(t_i + k\Delta t) - b(t_j + k\Delta t)) \approx 0.$$

As before, we have $\sum_{k=0}^{d-1} (b(t_i + k\Delta t) - b(t_j + k\Delta t))^2 \approx 0$. We conclude that

$$\|\mathbf{x}(t_i) - \mathbf{x}(t_j)\|^2 = \sum_{k=0}^{d-1} w^2(t_i + k\Delta t). \quad (5)$$

This sum measures the energy of the (sampled) seismic wave over the interval $[t_i, t_i + d\Delta t)$. Because the patch size, $d\Delta t$, is chosen so that $w(t)$ oscillates several times over the patch (see Fig. 3), the interval $[t_i, t_i + d\Delta t)$ is comprised of several wavelengths of $w(t)$, and the energy (5) is usually large.

Finally, we consider the case where both patches contain part of the seismic wave $w(t)$ (see Fig. 3),

$$\begin{aligned} \|\mathbf{x}(t_i) - \mathbf{x}(t_j)\|^2 &= \sum_{k=0}^{d-1} (b(t_i + k\Delta t) - b(t_j + k\Delta t))^2 + \sum_{k=0}^{d-1} (w(t_i + k\Delta t) - w(t_j + k\Delta t))^2 \\ &\quad + 2 \sum_{k=0}^{d-1} (w(t_i + k\Delta t) - w(t_j + k\Delta t)) (b(t_i + k\Delta t) - b(t_j + k\Delta t)). \end{aligned}$$

If we assume that the baseline signal varies slowly over time, then we have again

$$\|\mathbf{x}(t_i) - \mathbf{x}(t_j)\|^2 \approx \sum_{k=0}^{d-1} (w(t_i + k\Delta t) - w(t_j + k\Delta t))^2. \quad (6)$$

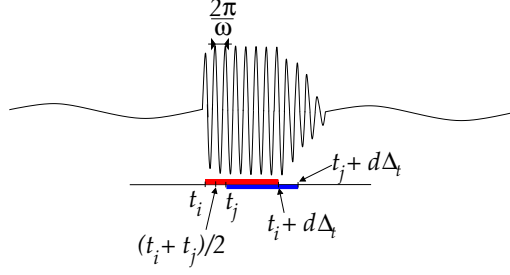


Figure 3: The two patches (in red and in blue) contain part of the seismic wave $w(t)$.

The sum (6) measures the energy of the difference between two overlapping sections of the seismic wave $w(t)$, sampled every Δt (see Fig. 3). In order to estimate the size of this energy, we approximate the seismic wave with a cosine function (see Fig. 3), $w(t) = \cos(\omega t)$, where the frequency ω corresponds to the peak of the short-time Fourier transform of $w(t)$ around τ . In this case, we have

$$\begin{aligned} w(t_i + k\Delta t) - w(t_j + k\Delta t) &= \cos\{\omega(t_i + k\Delta t)\} - \cos\{\omega(t_j + k\Delta t)\} \\ &= 2 \sin\{\omega(t_j - t_i)/2\} \sin\{\omega[(t_i + t_j)/2 + k\Delta t]\}, \end{aligned}$$

and the sum (6) becomes

$$\sum_{k=0}^{d-1} \{w(t_i + k\Delta t) - w(t_j + k\Delta t)\}^2 = 4 \sin^2\{\omega(t_j - t_i)/2\} \sum_{k=0}^{d-1} \sin^2\{\omega[(t_i + t_j)/2 + k\Delta t]\}. \quad (7)$$

The sum in the right-hand side of (7) can be written as

$$\sum_{k=0}^{d-1} \sin^2\{\omega[(t_i + t_j)/2 + k\Delta t]\} = \sum_{k=0}^{d-1} \cos^2\left\{\omega\left[(t_i + t_j)/2 + \frac{\pi}{2\omega} + k\Delta t\right]\right\} \quad (8)$$

The right-hand side of (8) is the energy of the cosine function sampled every Δt on an interval starting at $(t_i + t_j)/2 + \pi/2\omega$ of length $d\Delta t$. As explained above $d\Delta t \gg 2\pi/\omega$, and thus the energy (8) is measured over several wavelengths and is therefore large. Finally, given a patch starting at time t_i , all patches starting at time t_j , where $t_j = t_i + (q + 1/2)2\pi/\omega$, for some $q = 0, 1, \dots$ will satisfy $\sin^2(\omega(t_j - t_i)/2) = 1$. We conclude that given a patch starting at time t_i , there are many choices of t_j such that the sum in (7), and therefore the norm in (6), are large.

In summary, we expect the mutual distance between patches extracted from the baseline signal to be small, while the mutual distance between baseline and arrival patches will be large. Moreover, we also expect that two arrival patches will often be at a large distance of one another.

2.3. Global Alignment of the Patches

We have seen that, given a seismogram $x(t)$, the patch trajectory $\{x(t), t \geq 0\}$ is a one-dimensional curve in \mathbb{R}^d . We could imagine that this curve would be wildly scattered all over \mathbb{R}^d , exploring every part of the space. In fact, this is not the case: if the seismogram has bounded derivatives up to order s , then the patch trajectory will lie near the intersection of s hyperplanes. In other words, the one-dimensional curve does not explore the entire space, but stays inside a subspace of dimension $d - s$.

We first observe that we can compute numerical estimates of the first $s \leq d - 1$ derivatives from the d coordinates of the patch $x(t)$. We can therefore quantify the local regularity of the function $x(t)$ over the time window $[t, t + d\Delta t]$ from the knowledge of $x(t)$. This observation was at the origin of the first embedding theorems [17] that did not use the notion of time-delay coordinates – which is equivalent to our

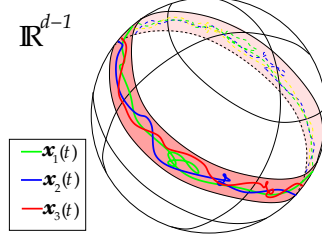


Figure 4: The patch trajectories $\mathbf{x}_1(t)$, $\mathbf{x}_2(t)$ and $\mathbf{x}_3(t)$ belong to a common low-dimensional submanifold of the $d-2$ -dimensional unit sphere in \mathbb{R}^{d-1} .

notion of patch – but involved the differential coordinates $d^p x/dt^p$, $p = 0, \dots, d-1$. We consider a forward finite difference scheme that approximates the derivative of order p ($1 \leq p \leq s-1$) at time t ,

$$\frac{d^p x}{dt^p} \approx \frac{1}{(\Delta t)^p} \sum_{j=0}^p a_j x(t + j\Delta t). \quad (9)$$

Obviously, more accurate schemes can be obtained using central differences; we use a forward scheme here to simplify the discussion. The finite difference (9) is a linear combination of the first $p+1$ coordinates of the patch $\mathbf{x}(t)$. We assume that $d^p x/dt^p$ is bounded by a small constant over an interval U . Consequently, the finite difference (9) is small, and we have

$$\left| \sum_{j=0}^p a_j x(t + j\Delta t) \right| \approx 0, \quad t \in U.$$

This last statement can be translated as follows: the distance of the patches $\{\mathbf{x}(t), t \in U\}$ to the hyperplane of \mathbb{R}^d defined by the vector

$$[a_0 \ a_1 \ \dots \ a_p \ 0 \ \dots \ 0]^T \in \mathbb{R}^d. \quad (10)$$

is small. We conclude that if the function $x(t)$ has s bounded derivatives for $t \in U$, then the trajectory of $\mathbf{x}(t)$, for $t \in U$, lies near the intersection of s hyperplanes, each of which is defined by a set of coefficients of the form (10).

2.4. Normalization of Patch-Space

We now consider the following question: if we want to use seismograms from different stations to learn the general shape of a seismic wave (see e.g. [18]), how should we normalize the seismograms? The magnitude of an earthquake, which characterizes its damaging effect, is defined as a logarithmic function of the radiated energy [19]. The radiated energy can be estimated by integrating the velocity associated with the displacement measured by seismograms [19]. A logarithmic normalization [19] would make it possible to account for the large variability in the energy and would allow us to compare seismograms from different stations or from different events. We favor an equivalent normalization that consists in rescaling each patch by its energy. For every patch $\mathbf{x}(t)$, we first remove any slowly varying drift by removing the mean of $x(t)$ over the interval $[t, t + d\Delta t)$, $\bar{x}(t) = \left(\sum_{k=0}^{d-1} x(t + k\Delta t) \right) / d$, and we compute the centered patch

$$\mathbf{x}_0(t) = [x(t) - \bar{x}(t) \ \dots \ x(t + (d-1)\Delta t) - \bar{x}(t)]^T.$$

Finally, we project the centered patch $\mathbf{x}_0(t)$ on the unit sphere and define the normalized patch

$$\frac{\mathbf{x}_0(t)}{\|\mathbf{x}_0(t)\|}. \quad (11)$$

The normalized patch (11) characterizes the local oscillation of $x(t)$ in a manner that is independent of changes in amplitude and of any slow drift of the seismogram. Geometrically, the trajectory of the normalized patch is a curve on the $d - 2$ dimensional unit sphere in \mathbb{R}^{d-1} (see Fig. 4). Indeed, after subtracting the mean $\bar{x}(t)$, the patch lies on the hyperplane of \mathbb{R}^d defined by $\sum_{i=1}^d x_i = 0$. After the normalization (11), the normalized patch lives at the intersection of the unit sphere in \mathbb{R}^d , and the hyperplane $\sum_{i=1}^d x_i = 0$. This intersection is also a unit sphere, but in \mathbb{R}^{d-1} . In the remaining of the paper we assume that each patch $\mathbf{x}(t)$ has been normalized according to (11).

2.5. The Embedding Dimension

The patch size is determined by the number, d , of time-delay coordinates. The selection of d can be guided by several algorithms that have been proposed in the context of the analysis of nonlinear dynamical systems from time series [20, 21, 22]. In practice, if d is chosen too large, then our understanding of the geometry of patch-space becomes restricted by the number of patches available. Indeed, the number of patches is limited by the number of time samples; but we need a number of patches that grows exponentially with the dimension, d , of patch-space to properly estimate the distribution of the patches [23].

3. A new parametrization of patch-space

3.1. From a single seismogram to several seismograms

Because we learn to detect arrivals using more than a single seismic trace (as in [18], for instance) we need to understand the structure of patch space when patches come from different seismograms acquired at different stations. After normalization, all the patches live on the $d - 2$ -dimensional unit sphere in \mathbb{R}^{d-1} , and are therefore characterized by $d - 2$ coordinates. Each seismogram gives rise to a one-dimensional trajectory on the unit sphere (see Fig. 4). We expect that the patch trajectories created by different seismograms will remain close to one another, and will not be spread across the unit sphere. Our experiments confirm that the trajectories belong to a m -dimensional submanifold embedded in the $d - 2$ -dimensional unit sphere, where $m \ll d - 2$.

3.2. From continuous to discrete patch-space

In practice, each seismogram is sampled with the sampling period Δt , and we obtain a time series

$$x_i = x(t_i), \quad \text{where } t_i = i\Delta t.$$

The uniform sampling (in time) of the seismogram leads to a non uniform sampling (in space) of the one-dimensional trajectory $\mathbf{x}(t)$. We define the discrete patch by

$$\mathbf{x}_i = [x_i \quad x_{i+1} \quad \dots \quad x_{i+(d-1)}]^T. \quad (12)$$

In order to identify patches that contain seismic waves, we want to characterize the m -dimensional submanifold of patch trajectories. Formally, we seek a smooth parametrization of the set of patch trajectories that uses the minimum number of parameters – theoretically only m . An answer to this question is provided by the computation of the principal components of the set of patches $\{\mathbf{x}_i, i = 0, \dots\}$. This method, known as singular-spectrum analysis [24, 25], has been used to characterize geophysical time series [26, 27]. Unfortunately, unless the patches lie along a low-dimensional linear subspace, many (typically more than m) principal components will be required to capture the curvature and torsion of the patch trajectories. This problem is quite severe since the part of patch-space that corresponds to arrivals exhibits high curvature. Our quest for a parametrization of the submanifold of patch trajectories is similar to the problem of reconstructing a phase space based on time-delay coordinates. Indeed, we can consider that seismograms are observables from the nonlinear dynamical system that is at the origin of the seismic waves. The authors in [28] and [29] show that patches extracted from volcanic tremors evolve around a low-dimensional attracting manifold (see also [30, 31, 32, 33]). The embedding dimension of the phase space reconstructed from Strombolian tremors was estimated in [28] to be around five (see also [29] and [34]). Our approach provides

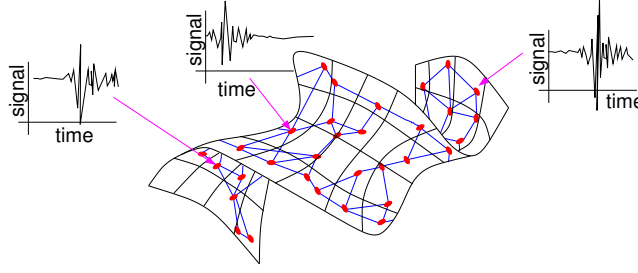


Figure 5: The patch graph: each node is a patch; nodes are connected if patches are similar.

a new perspective on this question by directly constructing a smooth parametrization of the set of phase spaces associated with the different seismograms. Our hypothesis, confirmed by our experiments, is that the vectors \mathbf{x}_i live close to a low-dimensional submanifold of the unit sphere in \mathbb{R}^{d-1} .

3.3. Nonlinear Parametrization of Patch-Space via the Eigenvectors of the Laplacian

3.3.1. The patch graph: a network of patches

Our plan is to assemble a global parametrization Ψ of the submanifold of patches from the knowledge of the pairwise distances between patches. In principle, we should measure the geodesic distances between patches along the underlying submanifold. Unfortunately, the only distances available to us are Euclidean distances. This issue can be resolved by observing that the Euclidean and the geodesic distances between two patches are very similar if the patches are in close proximity. We therefore use only distances between neighboring patches, and disregard distances between faraway patches.

We describe patch-space with a graph G that is constructed as follows. We assume that we have access to N patches $\mathbf{x}_i, i = 1, \dots, N$ extracted from several seismograms recorded at different stations. All the patches have been properly normalized, as explained in Sec. 2.4. Each patch \mathbf{x}_i becomes the vertex \mathbf{x}_i of the graph¹. Edges between vertices quantify the proximity between patches. Each vertex \mathbf{x}_i is connected to its ν_{NN} nearest neighbors according to the Euclidean distance $\|\mathbf{x}_i - \mathbf{x}_j\|$. When the vertices \mathbf{x}_i and \mathbf{x}_j are connected by an edge $\{i, j\}$ we write $\mathbf{x}_i \sim \mathbf{x}_j$. The weight $W_{i,j}$ on the edge $\{i, j\}$ measures the similarity between the patches \mathbf{x}_i and \mathbf{x}_j and is defined by

$$W_{i,j} = \begin{cases} e^{-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / \sigma^2}, & \text{if } \mathbf{x}_i \text{ is connected to } \mathbf{x}_j, \\ 0 & \text{otherwise.} \end{cases} \quad (13)$$

The scaling factor σ modulates our definition of proximity. If $\sigma \gg \max_{i,j} \|\mathbf{x}_i - \mathbf{x}_j\|$, then for all edges $\{i, j\}$ we have $W_{i,j} \approx 1$. This choice of σ blurs the distinction between baseline and arrival patches by pretending that all patches are similar to one another ($W_{i,j} \approx 1$). On the other hand, if $\sigma \approx 0$, then $W_{i,j} \approx 0$ for all edges $\{i, j\}$ such that $\|\mathbf{x}_i - \mathbf{x}_j\| > 0$. Only if $\|\mathbf{x}_i - \mathbf{x}_j\| \approx 0$ do we have $W_{i,j} \neq 0$. This choice of σ accentuates the differences between patches by pretending that all patches are different the minute they are slightly different. Obviously, this choice of σ is very sensitive to any noise existing in the data. The weighted graph G is fully characterized by the $N \times N$ weight matrix \mathbf{W} with entries $W_{i,j}$. We also define the diagonal degree matrix \mathbf{D} with entries $D_{i,i} = \sum_j W_{i,j}$.

3.3.2. We trust only local distances between patches

The construction of the parametrization of patch-space is guided by the following two principles:

- Distances between patches connected by an edge are small and one can approximate their geodesic distance by their Euclidean distance. This allows one to measure local geodesic distance without an existing knowledge of the underlying submanifold of patch trajectories.

¹We slightly abuse notation here: \mathbf{x}_i is patch as well as a vertex on the graph.

- The new parametrization of patch-space assembles the different constraints provided by the local geodesic distances into a set of global coordinates that vary smoothly over the submanifold.

Only an isometry will preserve exactly Euclidean distances between patches, and the isometry which is optimal for dimension reduction is given by PCA. However, as shown in the experiments, the coordinates provided by PCA are unable to capture the nonlinear structures formed by patch-space. We propose therefore to seek a sequence of functions ψ_1, ψ_2, \dots that will become the new coordinates of each patch.

3.3.3. The new coordinate functions: the eigenfunctions of the Laplacian

We define each coordinate function ψ_k as the solution to the following minimization problem

$$\min_{\psi_k} \frac{\sum_{\mathbf{x}_i \sim \mathbf{x}_j} W_{i,j} (\psi_k(\mathbf{x}_i) - \psi_k(\mathbf{x}_j))^2}{\sum_i D_{i,i} \psi_k^2(\mathbf{x}_i)}, \quad (14)$$

where ψ_k is orthogonal to the previous functions $\{\psi_0, \psi_1, \dots, \psi_{k-1}\}$,

$$\langle \psi_k, \psi_j \rangle = \sum_{i=1}^N D_{i,i} \psi_k(\mathbf{x}_i) \psi_j(\mathbf{x}_i) = 0 \quad (j = 0, \dots, k-1). \quad (15)$$

The numerator of the Rayleigh ratio (14) is a weighted sum of the gradient of ψ_k measured along the edges $\{i, j\}$ of the graph; it quantifies the average local distortion introduced by the map ψ_k . The distortion is measured locally: if \mathbf{x}_i and \mathbf{x}_j are far apart, then $W_{i,j} \approx 0$, and the difference $(\psi_k(\mathbf{x}_i) - \psi_k(\mathbf{x}_j))$ does not contribute to the sum (14). The denominator provides a natural normalization. The constraint of orthogonality (15) to the previous coordinate functions guarantees that the coordinates ψ_0, ψ_1, \dots describe the dataset with several resolutions: if $\langle \psi_k, \psi_j \rangle = 0$ then ψ_k experiences more oscillations on the dataset than the previous ψ_j . Intuitively, ψ_k plays the role of an additional digit that describes the location of \mathbf{x}_i with more precision. It turns out [35] that the solution of (14,15) is the solution to the generalized eigenvalue problem,

$$(\mathbf{D} - \mathbf{W}) \psi_k = \lambda_k \mathbf{D} \psi_k, \quad k = 0, 1, \dots \quad (16)$$

The first eigenvector ψ_0 , associated with $\lambda_0 = 0$, is constant, $\psi_0(\mathbf{x}_i) = 1, i = 1, \dots, N$; it is therefore not used. Finally, the new parametrization Ψ is defined by

$$\mathbf{x}_i \mapsto \Psi(\mathbf{x}_i) = [\psi_1(\mathbf{x}_i) \quad \psi_2(\mathbf{x}_i) \quad \dots \quad \psi_m(\mathbf{x}_i)]^T. \quad (17)$$

The matrix $\mathbf{P} = \mathbf{D}^{-1}\mathbf{W}$ is a row-stochastic matrix, associated with a Markov chain on the graph, and the matrix $\mathbf{L} = \mathbf{I} - \mathbf{P}$ that appears in (16) is known as the graph Laplacian. The idea of parametrizing a manifold using the eigenfunctions of the Laplacian can be traced back to ideas in spectral geometry [36], and has been developed extensively by several groups recently [37, 38, 39, 40]. The construction of the parametrization is summarized in Fig. 6 (see also [41] for a version that can be implemented with fast algorithms). Unlike PCA, which yields a set of vectors on which to project each \mathbf{x}_i , this nonlinear parametrization constructs the new coordinates of \mathbf{x}_i by concatenating the values of the $\psi_k, k = 1, \dots, m$ evaluated at \mathbf{x}_i , as defined in (17).

3.3.4. How many new coordinates do we need?

Our goal is to construct the most parsimonious parametrization of patch-space with the smallest number m of coordinates. We expect that if m is too small, then the new parametrization will not describe the data with enough precision, and the detection of seismic waves will be inaccurate. On the other hand, if m is too large, then some coordinates will be mainly describing the noise in the dataset (and not adding additional information), and the classification algorithm will overfit the training data. The experiments confirmed that $m = 25$ yields the optimal detection of seismic waves – and performed better than $m = 50$ – even when as many as $d = 1024$ time samples are included in each patch. Clearly, this approach results in a very significant reduction of dimensionality.

Algorithm 1: Parametrization of patch-space

Input: set of N patches, $\mathbf{x}_i, i = 1, \dots, N$; m : dimension of the embedding
 σ : width of the kernel for the graph; ν_{NN} : number of neighbors of each patch;

Algorithm:

1. construct the graph defined by the ν_{NN} nearest neighbors of each \mathbf{x}_i
2. compute the weight matrix \mathbf{W} , and the degree matrix \mathbf{D}
3. compute the m eigenvectors ψ_1, \dots, ψ_m of $\mathbf{D}^{-\frac{1}{2}} \mathbf{W} \mathbf{D}^{-\frac{1}{2}}$

Output: m coordinates $\Psi(\mathbf{x}_i) = [\psi_1(\mathbf{x}_i) \quad \psi_2(\mathbf{x}_i) \quad \dots \quad \psi_m(\mathbf{x}_i)]^T$.

Figure 6: Construction of the embedding

3.4. The case of a single station and a single event: a dynamical system connection

We consider in this section the simpler situation where patch space is constructed using only seismograms recorded at a single station from a single event. We identify patch space with the phase space, reconstructed by time-delay embedding, of the dynamical system associated with the earthquake. In the following, we explore the connections between the graph Laplacian defined on phase space and several methods that have been proposed to characterize a dynamical system.

3.4.1. Recurrence quantification analysis

We first observe that the weight matrix \mathbf{W} defined by (13) is closely related to the recurrence plot matrix associated with a dynamical system [42]. The recurrence plot matrix \mathbf{R} is defined by

$$R_{i,j} = \begin{cases} 1 & \text{if } \|\mathbf{x}_i - \mathbf{x}_j\| < \delta, \\ 0 & \text{otherwise.} \end{cases}$$

We can interpret the recurrence plot matrix \mathbf{R} in terms of the weight matrix of a graph defined as follows: each vertex \mathbf{x}_i is connected to the vertices that are within a distance δ of \mathbf{x}_i , and the weights along the edges are equal to one. This graph is similar to the ν_{NN} -nearest neighbor graph used in this work when δ is chosen so that on average each vertex is connected to ν_{NN} vertices, and $\sigma = \infty$ in (13). Several methods have been proposed to recover information about $\mathbf{x}(t)$ from the knowledge of \mathbf{R} [43, 44]. In this paper, we demonstrate that we can indeed construct a smooth parametrization of the phase space $\{\mathbf{x}(t), t \geq 0\}$ from the knowledge of the weight matrix \mathbf{W} only.

3.4.2. Complex networks in phase space

The recurrence plot matrix \mathbf{R} can also be interpreted as the adjacency matrix of a graph that captures the dynamical system. Several authors have proposed to characterize nonlinear dynamical systems by studying such graphs, known as “recurrence networks” [45, 46, 47] (see also [48, 49] for graphs that are constructed using the ν_{NN} nearest neighbors in phase space, as we do). After the graph is constructed, geometric properties of the graph, such as diameter, path lengths distribution, are computed (see [45] and references therein for a detailed review). Because such geometric properties can be computed directly from the eigenvalues λ_k defined by (16) [35], the methods proposed in [45, 46, 47, 48, 49] are in fact equivalent to the problem of studying the Laplacian defined on the network. Our approach can therefore be understood as a spectral characterization of phase space obtained from the eigenfunctions and the eigenvalues of the Laplacian defined on a recurrence network.

3.4.3. Frobénius-Perron Operator

Finally, one can define a Frobénius-Perron operator, similar to the matrix \mathbf{P} , that characterizes the evolution of the probability distribution of the configurations of phase space [50]. The eigenfunctions of the Frobénius-Perron operator can be used to partition phase space into almost invariant subsets [51, 52, 53]: a trajectory initiated in each of these subsets remains in the subset for a long time before escaping to another almost invariant subset [50]. The eigenfunctions of the Perron-Frobénius operator are exactly similar to the ψ_k . In summary, in the simpler case of the analysis of a single earthquake from a single station, our approach can be interpreted as a method to decompose phase space in terms of regions where the long term dynamics are similar [53].

4. Estimation of Arrival-Times of Seismic Waves

4.1. Learning the presence of seismic waves in patch-space

Our goal is to learn the association between the presence of a seismic wave within a patch, and the values of the patch coordinates. As explained before, we advocate a geometric approach: we expect that patches will organize themselves on the unit sphere in \mathbb{R}^{d-1} in a manner that will reveal the presence of seismic waves. We represent all the patches with the coordinates defined by Ψ in (17). We then use training data (labelled by experts) to partially populate patch-space with information about the presence or absence of seismic waves. We combine the information provided by the labels with the knowledge about the geometry of patch-space to train a classifier; this approach is known as semi-supervised [54]. We then use the classifier to classify unlabelled patches into baseline, or arrivals patches. The classification problem is formulated as a kernel ridge regression problem [55]: for any given patch, the classifier returns a number between 0 and 1 that quantifies the probability that a seismic wave be present within the patch.

We assume that N_l of the N patches have been labelled by an expert (analyst): for each of these patches we know if a seismic wave was observed in the patch, and at what time. We construct a response function f defined on the new coordinates, $\Psi(\mathbf{x}_i) \in \mathbb{R}^m$, and taking values in $[0, 1]$,

$$f : \mathbb{R}^m \longrightarrow [0, 1] \quad (18)$$

$$\Psi(\mathbf{x}_i) \longrightarrow f(\Psi(\mathbf{x}_i)). \quad (19)$$

The classifier decides that the patch \mathbf{x}_i contains an arrival if the response $f(\Psi(\mathbf{x}_i))$ is greater than some threshold $\varepsilon > 0$. The threshold ε controls the rates of false alarms and missed detections: a small ε results in many false alarms but will rarely miss arrivals, and vice versa. We expand the response function as a linear combination of Gaussian kernels in \mathbb{R}^m ,

$$f(\Psi(\mathbf{x})) = \sum_{j=1}^{N_l} \beta_j \exp \left\{ -\|\Psi(\mathbf{x}) - \Psi(\mathbf{x}_j)\|^2 / \alpha^2 \right\}. \quad (20)$$

The vector of unknown coefficients $\beta = [\beta_1, \dots, \beta_{N_l}]$ is computed using the training data. The kernel ridge regression [55] combines two ideas: distances between patches are measured using the Gaussian kernel \mathbf{K} , with entries $K_{i,j} = \exp \left\{ -\|\Psi(\mathbf{x}_j) - \Psi(\mathbf{x}_i)\|^2 / \alpha^2 \right\}$, $i, j = 1, \dots, N_l$; and the classifier is designed to provide the simplest model of the response in terms of the N_l training data. Rather than trying to find the optimal fit of the function f to the N_l labelled patches, we penalize the regression (20) by imposing a penalty on the norm of β [55]. This prevents the model (20) from overfitting the training samples. The optimal regression is defined as the solution to the quadratic minimization problem

$$\|\mathbf{r} - \mathbf{K}\beta\|^2 + \mu\beta^T \mathbf{K}\beta, \quad (21)$$

where $\mathbf{r} = [r_1 \dots r_{N_l}]^T$ is the known response on the N_l labelled patches. The parameter μ controls the amount of penalization: $\mu = 0$ yields a least squares fit, while $\mu = \infty$ ignores the data. For a given choice of μ , the optimal vector of coefficients [55] is given by

$$\beta = (\mathbf{K} + \mu\mathbf{I})^{-1} \mathbf{r}, \quad (22)$$

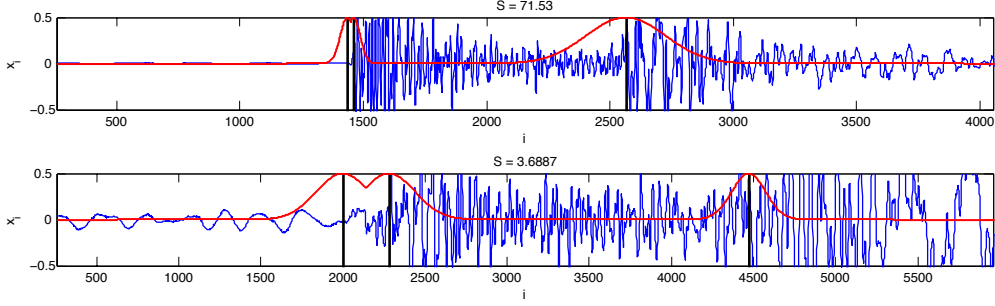


Figure 7: Seismic traces x_i (blue); estimated responses r_i (red); arrival-times τ_j (black vertical bars).

where \mathbf{I} is the $N_l \times N_l$ identity matrix. In our experiments, the ridge parameter μ was determined by cross-validation, and the same value, $\mu = 0.8$, was used throughout. The Gaussian width α is chosen to be a multiple of the average kernel distance,

$$\alpha^2 = C \left(\text{average}_{\mathbf{x}_i, \mathbf{x}_j} \|\Psi(\mathbf{x}_i) - \Psi(\mathbf{x}_j)\|^2 \right); \quad (23)$$

for all experiments we chose $C = 0.51$.

4.2. Defining Ground truth

4.2.1. Uncertainty in arrival-time

In order to validate our approach we need to compare the output of the response function f , defined by (20), to the actual decision provided by an expert (analyst). This comparison needs to be performed for every patch being analyzed. Unfortunately, the decision of the analyst is usually formulated as a binary response: an arrival is present at time τ_i or not. We claim that this apparent perfect determination of the arrival-time is misleading. Indeed, as was pointed out in [56], the origin time and the arrival-time at a given station are, for all practical purposes, random variables whose distributions depend on quality of the seismic record and the training and experience of the analyst detecting the arrivals. We formalize this intuition and model the arrival-time estimated by the analyst as a Gaussian distribution with mean τ_j and variance h_j . The parameter h_j controls the width of the Gaussian and quantifies the confidence with which the analyst estimated τ_j . Ideally, h_j should be a function of the inter-observer variability for the estimation of τ_j . In this work, we propose to estimate the uncertainty h_j directly from the seismogram. For each arrival-time τ_j , we compute the dominant frequency ω of $x(t)$ using a short Fourier transform. Let $T = 1/\omega$ be the period associated with ω , we define the uncertainty h_j as follows

$$h_j = \begin{cases} 2T/\Delta t & \text{if } 2T/\Delta t < h_{\max}, \\ h_{\max} & \text{otherwise.} \end{cases}$$

This choice of h_j corresponds to the following idea: if the seismogram were to be a pure sinusoidal function oscillating at the frequency ω (see Fig 3), then this choice of h_j would guarantee that we observe two periods (cycles) of $x(t)$ over a time interval of length h_j .

Finally, we define the true response r_i at time t_i to be the maximum of the Gaussian bumps associated with the arrival-time times τ_j nearest to time t_i ,

$$r_i = \max_j \left\{ \exp \left(-(t_i - \tau_j)^2 / h_j \right) \right\}. \quad (24)$$

Figure 7 displays two seismograms with different values for h_j . In the top seismogram the first arrival is very localized (small h_1), whereas the second arrival corresponds to a lower instantaneous frequency, and is therefore less localized (large h_2). In the second seismogram (bottom of Fig. 7) the first two arrivals are very close to one another resulting in an overlap of the Gaussians defining the response r_i .

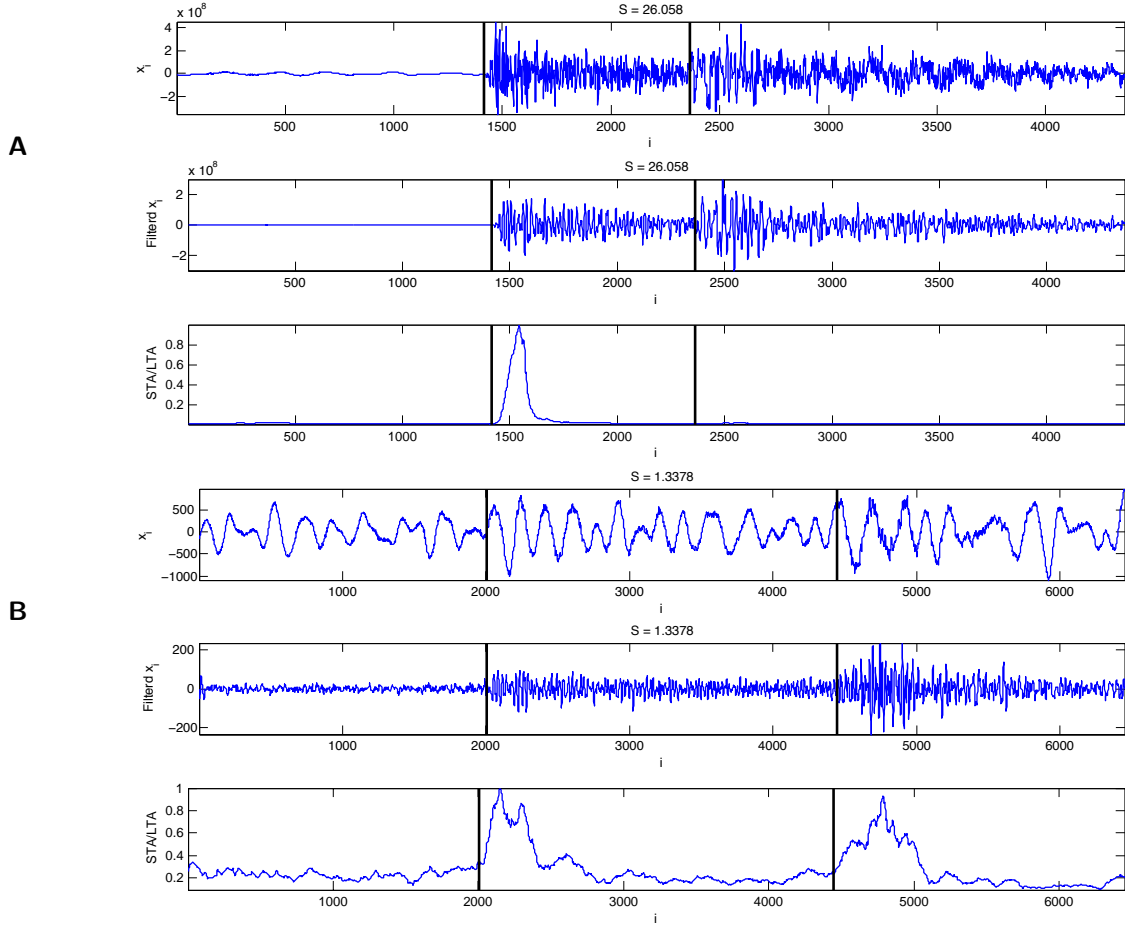


Figure 8: Raw and filtered seismic traces with associated STA/LTA outputs; high energy localization (**A**) and very diffuse energy localization (**B**). Analyst picks are represented by bars.

4.2.2. Energy localization of seismic traces

Because we analyze seismic traces of very different quality, we need to define a concept of energy localization associated with a seismic trace. Indeed, variability in the estimation of arrival-times in human is less pronounced when a seismic trace contains very localized arrivals [57]. We plan therefore to evaluate the performance of our method as a function of the energy localization level. For this purpose, we define the energy localization of a given trace to be the ratio of the energy of the seismic waves over the energy of the baseline activity. This ratio can be defined in terms of the set of patches that are extracted from the trace. For a given trace, let \mathcal{A} be the subset of patches that contain arrivals, and \mathcal{B} the complement of \mathcal{A} : the subset of patches that contain only baseline activity. We define the energy localization by the ratio

$$S = \frac{|\mathcal{B}|}{|\mathcal{A}|} \left(\sum_{\mathbf{x}_i \in \mathcal{A}} \|\mathbf{x}_i\|^2 \right) / \left(\sum_{\mathbf{x}_i \in \mathcal{B}} \|\mathbf{x}_i\|^2 \right), \quad (25)$$

where $|\mathcal{A}|$ (resp. $|\mathcal{B}|$) is the cardinality of \mathcal{A} (resp. \mathcal{B}). Figure 8 shows two seismic traces with very different energy localizations. Arrival-times assigned by the analyst are represented by vertical bars. The short window used to compute the STA is three second long (120 samples) and the long window, which immediately follows the short window, is 27 second long (1080 samples). Before computing the STA/LTA ratio, we apply a standard preprocessing step and filter the raw seismic trace with a passband ($[0.8, 3.5]$ Hz)

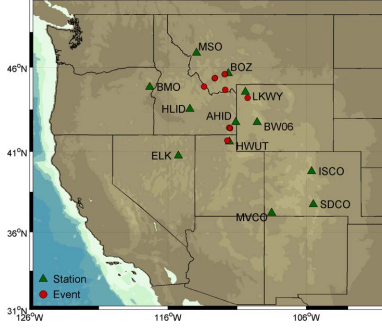


Figure 9: Locations of the stations and events from the Rocky Mountain region.

Butterworth filter (see Fig. 8). For the purpose of visual comparison, we normalized the STA/LTA output so that its maximum value is one. The trace **(A)** in Fig. 8 has a large energy localization, while the trace **(B)** has a very low energy localization. The Butterworth filter is able to remove some of the irrelevant low-frequency oscillations in **(B)** and yields a signal that can be processed by STA/LTA. We note that the second arrival (Lg) in the first trace **(A)** is missed by the STA/LTA algorithm.

5. Results

5.1. Rocky Mountain Dataset

We validate our approach with a dataset composed of broadband seismic traces from seismic events that occurred in Idaho, Montana, Wyoming, and Utah, between 2005 and 2006 (see Fig. 9). Arrival-times have been determined by an analyst. The ten events with the largest number of arrivals were selected for analysis. In total, we used 84 different station records from ten different events containing 226 labelled arrivals. Of the 226 labelled arrivals, there are 72 Pn arrivals, 70 Pg arrivals, 6 Sn arrivals, and 78 Lg arrivals. The sampling rate was $1/\Delta t = 40$ Hz. We consider only the vertical channel in our analysis. To minimize the computational cost, patches are spaced apart by $40\Delta t$ (one second).

5.2. Validation of the Classifier

The performance of the algorithm varies as a function of the energy localization S , and therefore we perform three independent validations by dividing the seismic traces into three homogeneous subsets: $n_1 = 27$ traces with low energy localization ($S < 3$), $n_2 = 29$ traces with medium energy localization ($3 \leq S \leq 18$), and the remaining $n_3 = 28$ traces with high energy localization ($S > 18$). For each subset s , ($s = 1, 2, 3$), we perform a standard leave-one-out cross-validation [55] using n_s folds as follows. We choose a test seismogram $x_t(t)$ among the n_s traces and compute the optimal set of weights (22) for the kernel ridge classifier (20) using the remaining $n_s - 1$ traces. Patches \mathbf{x}_i are then randomly selected from the test seismogram $x_t(t)$ and the classifier computes the response function $f(\Psi(\mathbf{x}_i))$. The response of the classifier is compared to the true response r_i for various false alarms and missed detections levels. We repeat this procedure for each possible test seismogram $x_t(t)$ among the n_s seismograms. Figure 10 details the cross-validation procedure. We quantify the performance of the classifier using a Receiver Operating Characteristic (ROC) curve [55]. The true detection rate (one minus the type II error) is plotted against the type I error (false alarm rate). We characterize each ROC curve by the area under the curve (the closer to one, the better).

5.3. Optimization of the parameters of the algorithm

The optimal values of the parameters were computed using cross-validation. This procedure turned out to be very robust, since we used the same parameters for all experiments. The optimal classification performance was achieved by choosing $\sigma = \infty$ and $\nu_{NN} = 32$ in the construction of the graph Laplacian. This is equivalent to setting the weights $W_{i,j}$ on the edges to be 1, and yields a graph that is extremely robust to noise. The influence of the patch size on the classification performance can be found in a series of

Algorithm 2: Cross validation of the classification

Input: Seismic traces, and the associated responses (r_i).

Algorithm:

```

for  $s = 1$  to 3                                     // for each subset of seismic traces
    extract a total of  $N$  patches from  $n_s$  distinct seismic traces.
    compute new coordinates  $\Psi(\mathbf{x}_i)$  of each patch  $\mathbf{x}_i$   $i = 1, \dots, N$ 
    for  $j = 1$  to  $n_s$                                    // evaluate the classifier for each seismic trace  $j$ 
        build classifier using all patches except those from trace  $j$ 
        for all patches  $\mathbf{x}_i^j$  in trace  $j$  compute classifier response  $f(\Psi(\mathbf{x}_i^j))$ 
        for  $\varepsilon = \varepsilon_l$  to  $\varepsilon_u$ 
            // populate the ROC curve using different thresholds to detect an arrival
            if  $f(\Psi(\mathbf{x}_i^j)) > \varepsilon$  and  $r_i < \varepsilon_0$  then declare false positive
            else if  $f(\Psi(\mathbf{x}_i^j)) < \varepsilon$  and  $r_i > \varepsilon_0$  then declare false negative
            end if
            record false positive and false negative rates for patches in fold  $j$ 
        end for
    end for
    compute average false positive and false negative rate

end for

Output: area under the ROC curve.

```

Figure 10: Cross validation procedure.

ROC curves in Fig. 12. We observe in Fig. 12 that the STA/LTA algorithm performs best for seismograms with low energy localization. Indeed, waveforms with high energy localization contain very little energy in the baseline part of the waveform. As a result, the STA/LTA ratio is much larger for the primary waves than it is for the secondary waves, and STA/LTA often misses the secondary waves. For traces with low energy localization, the ratio between baseline and arrivals energies is much smaller. This causes the STA/LTA ratio to reach similar values for the primary and secondary arrivals. Figure 13 shows three seismic traces with different energy localizations. STA/LTA (magenta) misses the secondary wave for medium and high energy localizations. Our approach (green) can detect the primary and secondary waves at all energy localization levels. For seismic traces with low energy localization our approach cannot compete with STA/LTA when d drops below 256. Of course, it is unfair to compare our approach using patches of only 128 samples with STA/LTA, which uses 1080 time samples (27 seconds). Indeed, as soon as $d \geq 1024$, our approach outperforms STA/LTA. Interestingly, our approach does not benefit from using a much larger d ; when $d = 2048$ the scale of the local analysis is no longer adapted to the physical process that we study.

5.4. So what does the set of patches look like?

To help us gain some insight into the geometric organization of patch space we display the patches using some of the new coordinates $\Psi(\mathbf{x}_i)$. For the three subsets of patches (classified according to the energy localization), we display in Fig. 11 each patch as a dot using three coordinates of the coordinate vectors Ψ . Because we can only display three coordinates, among the 25 (or 50) that we use to classify the patches, we use the three coordinates that best reveal the organization of patch-space. The color of the dot encodes the presence (orange) or absence (blue) of an arrival within \mathbf{x}_i . As the energy localization increases the separation between baseline patches and arrival patches improves. This visual impression is confirmed using the quantitative evaluation performed with the ROC curves (see Fig. 12). Clearly the shape of the set of patches is not linear, and would not be well approximated with a linear subspace.

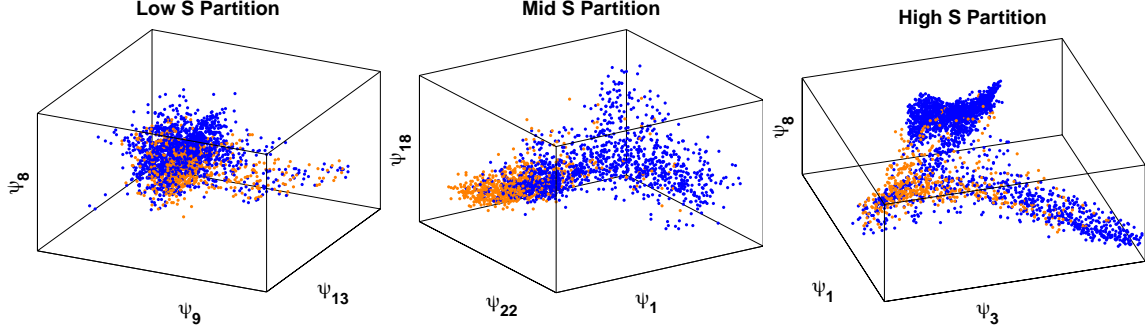


Figure 11: Each patch \mathbf{x}_i is represented as a dot using three coordinates of the vector Ψ . The color encodes the presence (orange) or absence (blue) of an arrival within \mathbf{x}_i . The energy localization levels increases from left to right.

5.5. Classification Performance

We first compare our approach to the gold-standard provided by STA/LTA. The second stage of the evaluation consists in quantifying the importance of the nonlinear dimension reduction Ψ defined by (17). To gauge the effect of Ψ we replace it by two linear transforms: a wavelet transform and a PCA transform. In both cases, we reduce the dimensionality of each patch from d to m . Wavelets have been used for a long time in seismology because seismograms can be approximated with very high precision using a small number of wavelet coefficients (e.g. [58, 10, 59] and references therein). On the other hand, we can also try to find the best linear approximation to a set of N patches. This linear approximation is obtained using PCA (also known as the singular-spectrum analysis [25], see section 3.2). The first m vectors of a PCA analysis yields the subspace that provides the optimal m -dimensional approximation to the set of patches. Our experiments demonstrate that the set of patches is not a linear structure and therefore is poorly approximated using PCA.

5.5.1. STA/LTA Ratio

We implement an STA/LTA detector as follows. We first apply a passband ([0.8 - 3.5] Hz) Butterworth filter to the raw traces. We then compute the ratio of the energy over two adjacent time window: a short window of 120 time samples (2 s) immediately followed by a long window of 1080 time samples (27 s). An arrival is detected when this ratio exceeds a threshold.

5.5.2. PCA and Wavelet Representations of Patch-Space

An orthonormal wavelet transform (symmlet 8) provides a multiscale decomposition of each patch \mathbf{x}_i in terms of d coefficients. Many of the coefficients are small and can be ignored. In order to decide which wavelet coefficients to retain, we select a fixed set of $m/2$ indices corresponding to the largest coefficients of the baseline patches. Similarly, we select the $m/2$ indices associated to the largest coefficients among the patches that contain arrivals. This procedure allows us to define a fixed set of m wavelet coefficients that are used for all patches as in input to the ridge regression algorithm. Similarly, we keep the first m coordinates returned by PCA.

5.5.3. Parameters of the classifier based on the PCA and Wavelet Representations

After applying a wavelet transform, or PCA, we use the same ridge classifier to detect arrivals. The parameters of the classifier are optimized for the wavelet and PCA transforms respectively. The Gaussian width α was again chosen to be a multiple of the average kernel distance between any two patches (see (23)). The parameter C in equation (23) was set to $C = 6.9$ for the wavelet parametrization, and $C = 4.6$ for the PCA parametrization. The ridge regression parameter was the same for both wavelets and PCA and was equal to $\mu = 10^{-3}$.

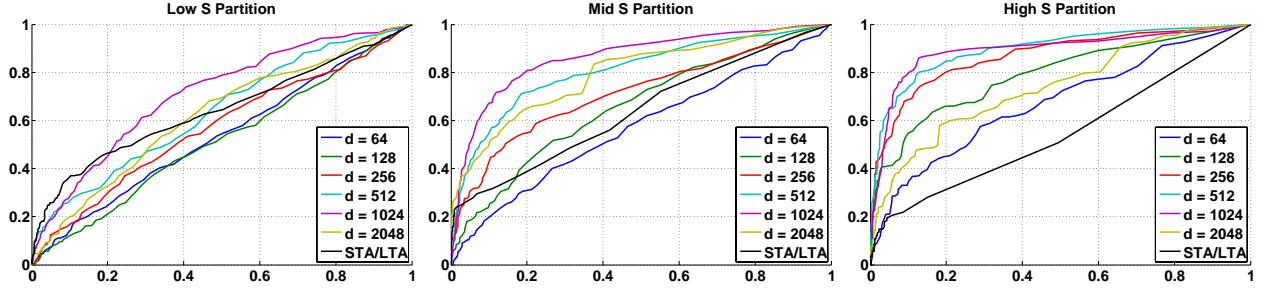


Figure 12: ROC curves for various values of the embedding dimension d at three levels of energy localization.

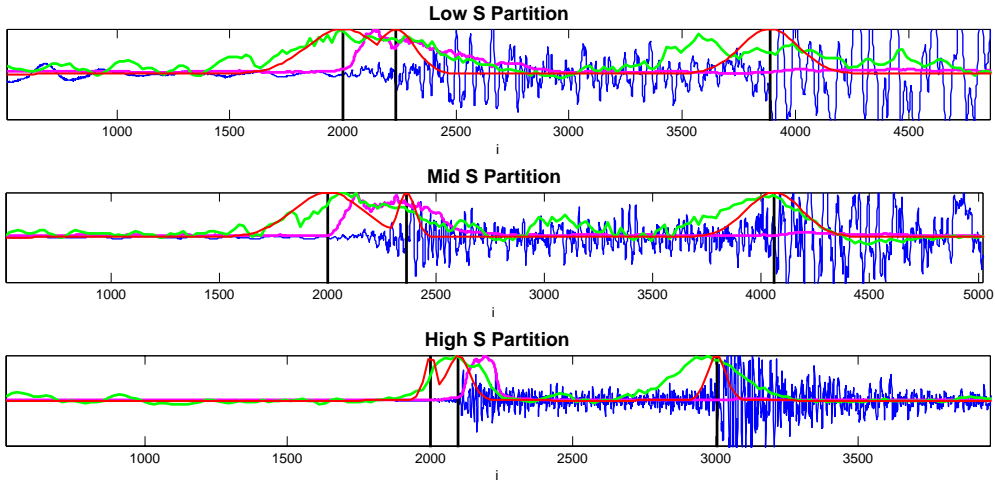


Figure 13: Seismic trace x_i (blue); true response r_i (red); STA/LTA (magenta); classifier $f(\Psi(x_i))$ (green).

6. Discussion

Table 1 provides a detailed summary of the performance of our approach. For each energy localization level (see section 5.2 for the definition of the subsets), we report the performance of the different detection methods as a function of d (patch dimension) and m (reduced dimension). The performance is quantified using the area under the ROC curve (the ROC curves are shown in Fig. 12); a perfect detector should have an area equal to one.

Effect of the patch size. As expected, the patch-based methods perform poorly if the patch is too small (there is not enough information to detect the seismic wave) or too large (the information is smeared over too large a window). The choice of the optimal patch size is dictated by the physical processes at stake here, since the optimal size is the same for all methods, irrespective of the transform used to reduce dimensionality. For high energy localization seismograms, the seismic waves are very localized and therefore all algorithms perform better with smaller patches (256 or 512 instead of 1024).

Effect of the transform used to reduce dimensionality. The experiments indicate that PCA outperforms a wavelet decomposition at every energy localization level. Both PCA and the wavelet transform are orthonormal transforms that can be understood in terms of a rotation of patch-space. PCA provides the optimal rotation to align patch-space along the m -dimensional subspace of best-fit. Finally, the nonlinear transformation Ψ based on the eigenfunctions of the Laplacian outperforms both PCA and wavelets. This

Table 1: Area under the ROC curve (closer to 1 is better) as a function of the patch dimension d and the reduced dimension m , at three different energy localization levels S .

	d	STA/LTA	Wavelet		PCA		Laplacian	
dimension			25	50	25	50	25	50
Low S	64	–	0.53	0.53	0.51	0.55	0.53	0.53
	128	–	0.55	0.49	0.52	0.51	0.52	0.52
	256	–	0.52	0.47	0.51	0.54	0.54	0.57
	512	–	0.53	0.50	0.61	0.61	0.62	0.64
	1024	0.66	0.43	0.38	0.54	0.64	0.70	0.71
	2048	–	0.45	0.39	0.55	0.48	0.61	0.62
Mid S	64	–	0.54	0.52	0.52	0.54	0.57	0.57
	128	–	0.57	0.55	0.53	0.56	0.66	0.66
	256	–	0.61	0.62	0.70	0.71	0.71	0.71
	512	–	0.68	0.67	0.76	0.79	0.79	0.81
	1024	0.68	0.77	0.76	0.81	0.84	0.86	0.86
	2048	–	0.64	0.66	0.69	0.75	0.80	0.80
High S	64	–	0.56	0.62	0.65	0.65	0.72	0.67
	128	–	0.68	0.69	0.78	0.73	0.80	0.79
	256	–	0.72	0.70	0.77	0.84	0.88	0.87
	512	–	0.74	0.79	0.73	0.85	0.90	0.90
	1024	0.59	0.72	0.76	0.67	0.75	0.88	0.89
	2048	–	0.51	0.49	0.57	0.67	0.76	0.74

clearly indicates that the set of patches contains nonlinear structures that cannot be well approximated by the optimal linear subspace computed by PCA. Interestingly, the results (not shown) are not improved by applying a wavelet transform before applying the nonlinear map Ψ (17) (see [60] for an example of a combination of wavelet transform with a nonlinear map similar to Ψ).

Dimension of patch-space. The performance is not significantly improved when 50 coordinates are used instead of 25. This is a result that is independent of the method used to reduced dimensionality, and is therefore a statement about the complexity of patch-space and about the physical nature of the seismic traces. As mentioned before, several studies have estimated the dimensionality of the low-dimensional inertial manifold reconstructed from the phase space of the tremors of a single volcano. This dimensionality was found in most studies [28, 29, 34] to be less than five: a number much smaller than our rough estimate of the dimensionality of patch space. Because patch space includes several seismograms from different events measured at different stations, we expect the dimensionality of this set to be greater than the dimensionality of the phase space reconstructed from the tremors of a single volcano measured at a single station. On the other hand, our study confirms that the combined phase spaces associated with regional seismic waves remains remarkably low-dimensional.

Computational complexity The complexity of this method is mainly determined by the combined complexity of the nearest neighbor search and the eigenvalue problem. We currently use the restarted Arnoldi method for sparse matrices implemented by the Matlab function `eigs` to solve the eigenvalue problem. We use the fast approximate nearest neighbor algorithm implemented by the ANN library [61] to construct the graph of patches.

Parametrization of slow manifolds with the eigenvectors of the Laplacian. As discussed in Sec. 3.2, our approach is related to the problem of parametrizing the low dimensional attracting manifold associated with the nonlinear dynamical system that is at the origin of the seismic waves. Similar ideas have been proposed in the context of molecular dynamics. In [39], the authors estimate slow variables that can be used to coarse-grain the dynamics near an inertial manifold. The slow variables are the eigenfunctions of a Fokker-Planck diffusion process defined on simulated data. Similarly, it was shown in [62, 51] that invariant

subsets associated with a dynamical system can be identified by studying the eigenvectors of a probability transition matrix – similar to the row-stochastic matrix $\mathbf{D}^{-1}\mathbf{W}$ defined in (16) – in the phase space of the dynamical system. Horenko [63] proposes to reconstruct the phase space generated by time-delay embedding using a combination of K of m -dimensional subspaces. In contrast, we use a nonlinear parametrization in this paper.

7. Concluding remarks

In this paper we presented a novel method to estimate from a seismogram arrival-times of seismic waves. We use time-delay embedding to characterize the local dynamics over temporal patches extracted from the seismic waveforms. We combine several delay-coordinates phase spaces formed by the different patch trajectories, and construct a graph that quantifies the distances between the different temporal patches. The eigenvectors of the Laplacian defined on the graph provide a low-dimensional parametrization of the combined phase spaces. Finally, a kernel ridge regression learns the association between each configuration of the phase space and the presence of a seismic wave. The regression is performed using the low-dimensional parametrization of the set of patches. Our approach outperforms standard linear techniques, such as wavelets and singular-spectrum analysis, and makes it possible to capture the nonlinear structures of the phase space reconstructed from time-delay embedding. This method unites the existing theory on time-delay embedding and the recent results on the nonlinear parametrization of manifold-valued data [37, 38, 39, 40]. We expect that our approach may be applicable to time series that are generated by other complex nonlinear dynamical processes, such as neurophysiological data, financial data, etc. We also expect that the idea of reconstructing phase space using the eigenvectors of the graph Laplacian may be used to remove noise from data generated by nonlinear dynamical systems [64, 65], and predict time-series [66, 67].

Acknowledgments

This work was supported by National Nuclear Security Administration Contract No. DE-AC04-94AL8500.

Author Contributions

Conception and design of methods and experiments: FGM and KMT. Acquisition of data: MJP and CJY. Analysis and interpretation of data: KMT, MJP, CJY, and FGM. Writing of manuscript: FGM and KMT.

References

- [1] R. Allen, Automatic phase pickers: Their present use and future prospects, *Bull. seism. Soc. Am.* 68 (1982) 1521–1532.
- [2] C. Panagiotakis, E. Kokinou, F. Vallianatos, Automatic P -Phase Picking Based on Local-Maxima Distribution, *IEEE Trans. Geosci. Remote Sens.* 46 (2008) 2280–2287.
- [3] R. Di Stefano, F. Aldersons, E. Kissling, P. Baccheschi, C. Chiarabba, D. Giardini, Automatic seismic phase picking and consistent observation error assessment: application to the Italian seismicity, *Geophys. J. Intern.* 165 (2006) 121–134.
- [4] W. Freiburger, An approximate method in signal detection, *Quarterly App. Math* 20 (1963) 373–378.
- [5] J. Berger, R. Sax, Seismic detectors: the state of the art, Technical Report, VELA Seismological Center, Alexandria, VA, 2001.
- [6] L. Persson, Statistical tests for regional seismic phase characterizations, *J. Seismol.* 7 (2003) 19–33.
- [7] C. Saragiotis, L. Hadjileontiadis, S. Panas, PAI-S/K: A robust automatic seismic P phase arrival identification scheme, *IEEE Trans. Geosci. Remote Sens.* 40 (2002) 1395–1404.
- [8] L. Küperkoch, T. Meier, J. Lee, W. Friederich, et al., Automated determination of P -phase arrival times at regional and local distances using higher order statistics, *Geophys. J. Intern.* 181 (2010) 1159–1170.
- [9] J. Galiana-Merino, J. Rosa-Herranz, S. Parolai, Seismic P Phase Picking Using a Kurtosis-Based Criterion in the Stationary Wavelet Domain, *IEEE Trans. Geosci. Remote Sens.* 46 (2008) 3815–3826.
- [10] H. Zhang, C. Thurber, C. Rowe, Automatic P -wave arrival detection and picking with multiscale wavelet analysis for single-component recordings, *Bull. seism. Soc. Am.* 93 (2003) 1904.
- [11] T. Bardainne, P. Gaillot, N. Dubos-Sallée, J. Blanco, G. Sénéchal, Characterization of seismic waveforms and classification of seismic events using chirplet atomic decomposition, *Geophys. J. Intern.* 166 (2006) 699–718.
- [12] M. Withers, R. Aster, C. Young, J. Beiriger, M. Harris, S. Moore, J. Trujillo, A comparison of select trigger algorithms for automated global seismic phase and event detection, *Bull. seism. Soc. Am.* 88(1) (1998) 95–106.

- [13] J. Wang, Adaptive training of neural networks for automatic seismic phase identification, *Pure appl. Geophys.* 159 (2002) 1021–1041.
- [14] T. Sauer, J. Yorke, M. Casdagli, Embedology, *J. Stat. Phys.* 65 (1991) 579–616.
- [15] H. Abarbanel, R. Brown, J. Sidorowich, L. Tsimring, The analysis of observed chaotic data in physical systems, *Rev. Modern Phys.* 65 (1993) 1331–1392.
- [16] R. Gilmore, Topological analysis of chaotic dynamical systems, *Rev. Modern Phys.* 70 (1998) 1455–1529.
- [17] N. Packard, J. Crutchfield, J. Farmer, R. Shaw, Geometry from a time series, *Phys. Rev. Lett.* 45 (1980) 712–716.
- [18] R. Aster, C. Rowe, Automatic phase pick refinement and similar event association in large seismic datasets, in: *Advances in seismic event location*, volume 18, pp. 231–263.
- [19] Y. Ben-Zion, Collective behavior of earthquakes and faults: Continuum-discrete transitions, progressive evolutionary changes, and different dynamic regimes, *Rev. Geophys.* 46 (2008).
- [20] K. Judd, A. Mees, Embedding as a modeling problem, *Physica D* 120 (1998) 273–286.
- [21] M. Kennel, H. Abarbanel, False neighbors and false strands: A reliable minimum embedding dimension algorithm, *Phys. Rev. E* 66 (2002) 26209.
- [22] M. Small, C. Tse, Optimal embedding parameters: a modelling paradigm, *Physica D* 194 (2004) 283–296.
- [23] D. Scott, *Multivariate Density Estimation*, Wiley, 1992.
- [24] D. Broomhead, G. King, Extracting qualitative dynamics from experimental data, *Physica D* 20 (1986) 217–236.
- [25] R. Vautard, M. Ghil, Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, *Physica D* 35 (1989) 395–424.
- [26] A. Sharma, D. Vassiliadis, K. Papadopoulos, Reconstruction of low-dimensional magnetospheric dynamics by singular spectrum analysis, *Geophys. Res. Lett.* 20 (1993) 335–338.
- [27] S. Gámiz-Fortis, D. Pozo-Vázquez, M. Esteban-Parra, Y. Castro-Díez, Spectral characteristics and predictability of the NAO assessed through Singular Spectral Analysis, *J. Geophys. Res.* 107 (2002) 4685.
- [28] E. De Lauro, S. De Martino, E. Del Pezzo, M. Falanga, M. Palo, R. Scarpa, Model for high-frequency Strombolian tremor inferred by wavefield decomposition and reconstruction of asymptotic dynamics, *J. Geophys. Res.* 113 (2008) B02302.
- [29] S. De Martino, M. Falanga, C. Godano, Dynamical similarity of explosions at Stromboli volcano, *Geophys. J. Intern.* 157 (2004) 1247–1254.
- [30] B. Chouet, H. Shaw, Fractal properties of tremor and gas piston events observed at Kilauea Volcano Hawaii, *J. Geophys. Res.* 96 (1991) 10177–10189.
- [31] C. Godano, C. Cardaci, E. Privitera, Intermittent behaviour of volcanic tremor at Mt. Etna, *Pure appl. Geophys.* 147 (1996) 729–744.
- [32] G. Yuan, M. Lozier, L. Pratt, C. Jones, K. Helfrich, Estimating the predictability of an oceanic time series using linear and nonlinear methods, *J. Geophys. Res.* 109 (2004).
- [33] K. Konstantinou, V. Schlindwein, Nature, wavefield properties and source mechanism of volcanic tremor: a review, *J. Volcanol. Geoth. Res.* 119 (2002) 161–187.
- [34] K. Konstantinou, Deterministic non-linear source processes of volcanic tremor signals accompanying the 1996 Vatnajökull eruption, central Iceland, *Geophys. J. Intern.* 148 (2002) 663–675.
- [35] F. Chung, *Spectral Graph Theory*, CBNS-AMS, 1997.
- [36] P. Bérard, G. Besson, S. Gallot, Embeddings Riemannian manifolds by their heat kernel, *Geom. Funct. Anal.* 4 (1994) 373–398.
- [37] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, *Neural Comput.* 15 (2003) 1373–1396.
- [38] R. R. Coifman, M. Maggioni, Diffusion wavelets, *Appl. Comput. Harmon. A.* 21 (2006) 53 – 94.
- [39] R. R. Coifman, I. G. Kevrekidis, S. Lafon, M. Maggioni, B. Nadler, Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems, *Multiscale Model. Sim.* 7 (2008) 842–864.
- [40] P. W. Jones, M. Maggioni, R. Schul, Manifold parametrizations by eigenfunctions of the Laplacian and heat kernels, *P. Natl. Acad. Sci. USA* 105 (2008) 1803–1808.
- [41] N. Saito, Data analysis and representation on a general domain using eigenfunctions of laplacian, *Appl. Comput. Harmon. A.* 25 (2008) 68–97.
- [42] J. Eckmann, S. Oliffson Kamphorst, D. Ruelle, Recurrence Plots of Dynamical Systems, *Europhys. Lett.* 4 (1987) 973–977.
- [43] N. Marwan, M. Carmen Romano, M. Thiel, J. Kurths, Recurrence plots for the analysis of complex systems, *Phys. Rep.* 438 (2007) 237–329.
- [44] G. Robinson, M. Thiel, Recurrences determine the dynamics, *Chaos* 19 (2009) 023104.
- [45] R. Donner, Y. Zou, J. Donges, N. Marwan, J. Kurths, Recurrence networks – A novel paradigm for nonlinear time series analysis, *New J. Phys.* 12 (2010) 033025.
- [46] Z. Gao, N. Jin, Complex network from time series based on phase space reconstruction, *Chaos* 19 (2009) 033137.
- [47] N. Marwan, J. Donges, Y. Zou, R. Donner, J. Kurths, Complex network approach for recurrence analysis of time series, *Phys. Lett. A* 373 (2009) 4246–4254.
- [48] Y. Shimada, T. Kimura, T. Ikeguchi, Analysis of Chaotic Dynamics Using Measures of the Complex Network Theory, *Artificial Neural Networks-ICANN 2008* (2008) 61–70.
- [49] X. Xu, J. Zhang, M. Small, Superfamily phenomena and motifs of networks induced from time series, *P. Natl. Acad. Sci. USA* 105 (2008) 19601.
- [50] P. Collet, J. Eckmann, *Concepts and results in chaotic dynamics: a short course*, Springer Verlag, 2006.
- [51] M. Dellnitz, M. Molo, P. Metzner, R. Preis, C. Schütte, Graph algorithms for dynamical systems, in: *Analysis, modeling and simulation of multiscale problems*, Springer, 2006, pp. 619–645.

- [52] M. Dellnitz, O. Junge, On the approximation of complicated dynamical behavior, *SIAM Journal on Numerical Analysis* 36 (1999) 491–515.
- [53] G. Froyland, M. Dellnitz, Detecting and locating near-optimal almost-invariant sets and cycles, *SIAM J. Sci. Comput.* 24 (2003) 1839–1863.
- [54] O. Chapelle, B. Schölkopf, A. Zien (Eds.), *Semi-Supervised Learning*, MIT Press, Cambridge, MA, 2006.
- [55] T. Hastie, R. Tibshinari, J. Freedman, *The elements of statistical learning*, Springer Verlag, 2009.
- [56] H. Freedman, “The Little Variable Factor” A Statistical Discussion of the Reading of Seismograms, *Bull. seism. Soc. Am.* 56(2) (1966) 593–604.
- [57] A. Velasco, C. Young, D. Anderson, Uncertainty in Phase Arrival Time Picks for Regional Seismic Events: An Experimental Design, Technical Report, US Department of Energy, 2001.
- [58] K. Anant, F. Dowla, Wavelet transform methods for phase identification in three-component seismograms, *Bull. seism. Soc. Am.* 87 (1997) 1598.
- [59] P. Gendron, J. Ebel, D. Manolakis, Rapid joint detection and classification with wavelet bases via Bayes theorem, *Bull. seism. Soc. Am.* 90 (2000) 764.
- [60] A. Schclar, A. Averbuch, N. Rabin, V. Zheludev, K. Hochman, A diffusion framework for detection of moving vehicles, *Digit. Signal Process.* 20 (2010) 111–122.
- [61] S. Arya, D. Mount, N. Netanyahu, R. Silverman, A. Wu, An optimal algorithm for approximate nearest neighbor searching fixed dimensions, *J. ACM* 45 (1998) 891–923. (available online at <http://www.cs.umd.edu/~mount/ANN/>).
- [62] P. Deuffhard, W. Huisinga, A. Fischer, C. Schütte, Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains, *Linear Algebra Appl.* 315 (2000) 39–59.
- [63] I. Horenko, On simultaneous data-based dimension reduction and hidden phase identification, *J. Atmos. Sci.* 65 (2008) 1941–1954.
- [64] D. Broomhead, J. Huke, M. Potts, Cancelling deterministic noise by constructing nonlinear inverses to linear filters, *Physica D* 89 (1996) 439–458.
- [65] E. Kostelich, T. Schreiber, Noise reduction in chaotic time-series data: a survey of common methods, *Phys. Rev. E* 48 (1993) 1752–1763.
- [66] M. Casdagli, Nonlinear prediction of chaotic time series, *Physica D* 35 (1989) 335–356.
- [67] H. Kantz, T. Schreiber, *Nonlinear time series analysis*, Cambridge University Press, 2004.